УДК 577.3

КРИТЕРИЙ ОЦЕНКИ СОСТОЯНИЯ СЛОЖНЫХ БИОСИСТЕМ

В.А. Фокин

Сибирский государственный медицинский университет
E-mail: fokin@ssmu.tomsk.ru

Предложен критерий интегральной оценки состояния биосистем, характеризующихся многомерными массивами данных. Критерий базируется на оценке внутримножественных расстояний в метрике Махаланобиса, что позволяет эффективно учесть взаимозависимость и внутри- и междуиндивидуальную вариабельность измеряемых показателей.

Формирование интегральных оценок состояния сложных биомедицинских систем представляет собой активно развивающееся направление медицинских информационных технологий [1, 2]. К основным свойствам биомедицинских данных, учет которых необходим при интегральной оценке состояния таких систем, следует отнести, прежде всего, статистическую взаимозависимость показателей между собой и внутри- и междуиндивидуальную вариабельность измеряемых показателей. Они являются проявлением системных свойств объекта исследования, отражением разнообразия системных реакций, формирующих конечное состояние исследуемой системы объекта, и характеризуют широту адаптивно-приспособительных возможностей объекта исследования

Сведение задач оценки состояния к выявлению одного или нескольких показателей, минимально коррелирующих между собой и максимально информативных по отношению к разнообразию исследуемых состояний системы, позволяет, во многих случаях, получать хорошее прагматическое решение задачи. Однако при этом существенно снижаются возможности познавательного исследования состояния системы, т.к. происходит уменьшение информации содержащейся в исходном множестве переменных о внутрисистемных взаимодействиях, формирующих результирующее состояние. В то же время, учет взаимосвязи измеряемых показателей может приводить к получению эффективных интегральных оценок даже в случаях малых изменений параметров системы самих по себе или при небольших уровнях внешних воздействиях [3], когда каждый из измеряемых показателей по отдельности может и не выходить за пределы среднестатистических норм.

Сложную систему достаточно трудно, а в большинстве случаев и невозможно охарактеризовать каким либо отдельным, экспериментально измеримым показателем, позволяющим оценивать ее состояние. Как правило, слабая формализация исследуемой ситуации требует проведения многомерных сравнений, что приводит к значительным трудностям в сопоставлении и интерпретации наблюдаемых результатов. В этих условиях анализ всего комплекса измеряемых в совокупности характеристик позволит адекватно оценивать происходящие в биосистеме изменения, а полученные на его основе интегральные оценки могут являться информативными количественными характеристиками ее состояния. В математической формулировке задача оценки состояния системы сводится к отображению пространства признаков, характеризующих систему, в одномерное пространство оценок состояний этой системы, определяемых величиной интегрального критерия.

Общий вид интегрального критерия

Будем оценивать состояние объекта $\vec{x}$, заданного набором признаков $(x_1,x_2,...,x_m)$, по отношению к некоторому референтному состоянию S_0, характеризуемому множеством объектов $\{\vec{x}_i\}$, $i=\overline{1,N_{S_0}}$, где N_{S_0} – количество объектов. В частности, в качестве референтного состояния в биомедицинских задачах может быть выбрано состояние здорового организма. Формирование множества объектов, представляющих референтное состояние, является плохоформализуемой задачей и требует, как правило, привлечения экспертных знаний специалистов соответствующей предметной области.

Состояние S_0 может быть отображено точками, занимающими некоторую область в соответствующем пространстве признаков. Конфигурацию этой области, взаимное расположение объектов в ней необходимо учитывать при проведении количественных оценок состояния. На рис. 1 иллюстрируется ситуация, когда неучет взаиморасположения объектов в областях, соответствующих состояниям S_1, S_2 и S_3, будет приводить к одинаковой оценке близости объекта $\vec{x}$ к каждому из них, если оценивать состояние объекта, как величину расстояния от объекта до эталонного представителя соответствующего множества.

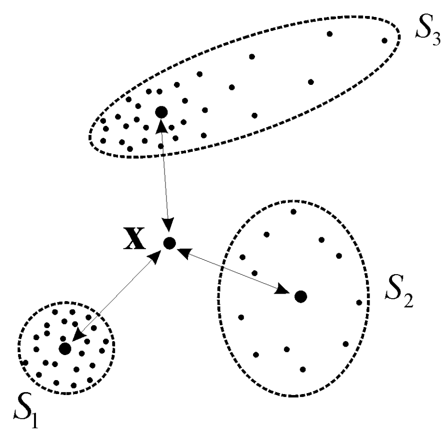


Рис. 1. *Одинаковое расстояние от объекта **X** до центров областей, характеризующихся различным взаиморасположением объектов*

Зададим интегральный критерий $I_{S_0}(\vec{x})$ оценки состояния объекта $\vec{x}$, по отношению к референтному состоянию S_0, следующим образом:

$$I_{S_0}(\vec{x}) = \frac{D(\vec{x}, S_0)}{D_{S_0}}, \tag{1}$$

где $D(\vec{x},S_0)$ – мера близости объекта $\vec{x}$ к состоянию S_0, D_{S_0} – мера компактности области, занимаемой объектами, относящимися к состоянию S_0, в соответствующем пространстве признаков. Количественно величина $D(\vec{x},S_0)$ может быть определена различным образом, зависящим как вида выбранного расстояния, так и от способа измерения расстояния от объекта до множества. Нормировка на величину D_{S_0} в выражении (1), характеризующая меру компактности области, соответствующую состоянию S_0, позволяет учесть вклад в интегральную оценку состояния конфигурации области и взаимного расположения объектов в ней.

Будем оценивать меру близости объекта $\vec{x}$ к состоянию S_0, как усредненное расстояние от него до всех объектов относящихся к данному состоянию

$$D(\vec{x}, S_0) = \frac{1}{N_{S_0}} \sum_{i=1}^{N_{S_0}} d(\vec{x}, \vec{x}_i).$$

Здесь через $d(\vec{x},\vec{x}_i)$ обозначено расстояние между объектами $\vec{x}$ и $\vec{x}_i$. Меру компактности D_{S_0} области, соответствующей состоянию S_0, зададим как усредненное расстояние средних расстояний от каждого объекта до всех оставшихся объектов:

$$D_{S_0} = \frac{1}{N_{S_0}} \sum_{i=1}^{N_{S_0}} \frac{1}{N_{S_0}-1} \sum_{j=1}^{N_{S_0}-1} d(\vec{x}_i, \vec{x}_j). \tag{2}$$

Без уменьшения общности сумму по j в выражении (2) можно распространить до N_{S_0}, т.к. соответствующее слагаемое $d(\vec{x}_j,\vec{x}_j)$, оценивающее расстояние от объекта до самого себя, будет равно нулю. Определенная таким образом величина D_{S_0} представляет собой внутримножественное расстояние и, в случае евклидового расстояния, равна удвоенной сумме дисперсий признаков [4]:

$$D_{S_0} = 2\sum_{k=1}^{m} \sigma_k^2,$$

где σ_k^2 – дисперсия k-ого признака, m – размерность пространства признаков.

Учет взаимозависимости признаков

Взаимосвязь признаков и вариабельность их значений может быть учтена путем оценки величины ковариации признаков, характеризующих референтное состояние S_0. В биомедицинских системах для оценки состояния наиболее эффективно использование расстояния Махаланобиса [2], при расчете которого используется ковариационная матрица. Расстояние Махаланобиса $d_M(\vec{x}_i,\vec{x}_j)$ между двумя объектами определяется следующим образом:

$$d_M(\vec{x}_i, \vec{x}_j) = (\vec{x}_i - \vec{x}_j)^T \mathbf{C}^{-1} (\vec{x}_i - \vec{x}_j).$$

Здесь $\vec{x}_i$ и $\vec{x}_j$ – векторы признаков i-ого и j-ого объектов, между которыми вычисляется расстояние, а $\mathbf{C}$ – матрица ковариации.

Оценим, что будет представлять собой D_{S_0} в случае расстояния Махаланобиса. Для этого рассмотрим ортонормированное линейное преобразование исходного пространства признаков $\vec{x}^* = \mathbf{A}\vec{x}$. Известно, что подобное преобразование пространства сохраняет взаимные расстояния между объектами и, следовательно, $D_{S_0}=D^*_{S_0}$. Выберем в качестве матрицы преобразования $\mathbf{A}$, матрицу следующего вида:

$$\mathbf{A} = \begin{pmatrix} \vec{e}_1^T \\ \vec{e}_2^T \\ \vdots \\ \vec{e}_m^T \end{pmatrix}.$$

Ее элементами являются компоненты собственных векторов $\vec{e}_k$ ковариационной матрицы $\mathbf{C}_0$:

$$\mathbf{C_0}\vec{e}_k = \lambda_k \vec{e}_k, \qquad k = \overline{1, m},$$

где λ_k – собственные значения.

Собственные векторы $\vec{e}_k$ образуют ортонормированное множество $\vec{e}_k^T\vec{e}_l=\delta_{kl}$, поэтому для матрицы $\mathbf{A}$ будут выполняться следующие соотношения:

$$\mathbf{AA^T} = \mathbf{A^TA} = \mathbf{I}, \qquad \mathbf{A} = \mathbf{A}^{-1} \tag{3}$$

и она преобразует ковариационную матрицу $\mathbf{C}_0$ в диагональную, элементами которой являются несмещенные оценки выборочной дисперсии:

$$\mathbf{C_0^*} = \mathbf{AC_0A^T} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_m \end{pmatrix} = \Lambda, \quad \lambda_k = \sigma_k^2.$$

Запишем внутримножественное расстояние (2) в преобразованном пространстве признаков:

$$D_{S_0}^* = \frac{1}{N_{S_0}(N_{S_0}-1)} \sum_{i=1}^{N_{S_0}} \sum_{j=1}^{N_{S_0}} (\vec{x}_i^* - \vec{x}_j^*)^T \Lambda^{-1} (\vec{x}_i^* - \vec{x}_j^*).$$

Поскольку матрица Λ является диагональной, то обратная к ней матрица Λ^{-1} также будет иметь диагональный вид, а элементы, расположенные по главной диагонали, будут равны $1/\lambda_k$. Учитывая соотношения (3) и переходя к координатной форме записи, получим следующее выражение для внутримножественного расстояния:

$$D_{S_0}^* = \frac{1}{N_{S_0}(N_{S_0}-1)} \sum_{i=1}^{N_{S_0}} \sum_{j=1}^{N_{S_0}} \sum_{k=1}^{m} \frac{(x_{i,k} - x_{j,k})(x_{i,k} - x_{j\,k})}{\sigma_k^2}, \tag{4}$$

где $x_{i,k}$ – значение k-ого признака i-ого объекта.

После перегруппировки слагаемых и ряда очевидных преобразований выражение (4) можно привести к следующему виду:

$$D^*_{S_0} = \frac{N_{S_0}}{(N_{S_0}-1)} \sum_{k=1}^{m} \frac{1}{\sigma_k^2} \left[\frac{1}{N_{S_0}^2} \sum_{i=1}^{N_{S_0}} \sum_{j=1}^{N_{S_0}} (x_{i,k} - x_{j,k})^2 \right] =$$

$$= 2\sum_{k=1}^{m} \frac{1}{\sigma_k^2} \left[\frac{N_{S_0}}{N_{S_0}-1} (\overline{x_k^2} - \overline{x}_k^{\,2}) \right].$$

Выражение, стоящее в квадратных скобках, представляет собой исправленную дисперсию k-ого признака σ_k^2. С учетом этого, количественная оценка меры компактности области, характеризующей референтное состояние S_0 в метрике Махаланобиса, будет равна удвоенной размерности пространства признаков

$$D_{S_0} = D^*_{S_0} = 2m.$$

Таким образом, критерий количественной оценки состояния некоторого объекта $\vec{x}$ может быть представлен в следующем виде:

$$I_{S_0}(\vec{x}) = \frac{1}{2mN_{S_0}} \sum_{i=1}^{N_{S_0}} d_M(\vec{x}, \vec{x}_i).$$

Выделение показателей, по совокупности которых следует производить интегральную оценку состояния биосистем, трудно формализуемая задача. Ее решение определяется целями проводимых исследований, ограничениями, накладываемыми на условия их проведения, используемой измерительной аппаратурой, уровнем знаний об объекте исследования и т.д. Вместе с тем, можно определить общий подход к их формированию, следующий из системных свойств данных. Он следует из того, что описание свойств любой системы укладывается в определенную иерархическую структуру [5], каждый уровень которой соотносится с соответствующими методами получения данных. Как правило, на уровне популяций и организмов используются эпидемиологические, социологические, психологические методы; на уровне органов или тканей применяются физиологические; на клеточном уровне – микрометрические, цитогенетические; на внутриклеточном уровне – биохимические и биофизические методы измерений и т.п. Поэтому комплекс показателей, который в совокупности отражает свойства соответствующего элемента структурного описания, должен учитываться при его интегральной оценке.

СПИСОК ЛИТЕРАТУРЫ

1. Миронкина Ю.Н., Бобров А.Ф. Информационная технология статистического синтеза критериев и алгоритмов оценки функционального состояния человека в прикладных медико-биологических исследованиях // Информационные технологии. – 1998. – № 3. – С. 41–47.
2. Генкин А.А. Новая информационная технология анализа медицинских данных (программный комплекс ОМИС). – СПб.: Политехника, 1999. – 191 с.
3. Конрадов А.А. Статистические подходы к анализу многомерных гетерогенных биологических систем // Радиационная биология, радиоэкология. – 1994. – Т. 34. – Вып. 6. – С. 877–886.
4. Ту Дж., Гонсалес Р. Принципы распознавания образов. – М.: Мир, 1978. – 416 с.
5. Клир Дж. Системология. Автоматизация решения системных задач / Пер. с англ. – М.: Радио и связь, 1990. – 554 с.

УДК 612.014.461